

Is rifabutin prophylaxis against *Mycobacterium avium* complex infection in HIV infection worthwhile? The net impact on patients suggests not

Judith M Stephenson, Ian G Williams

Introduction

The purpose of this article is to examine the validity and relevance of studies of medical treatment, using rifabutin prophylaxis as an example. Excellent guides to assessing medical literature have been described in some detail by Guyatt and others.^{1,2} This article draws on those guides to address the question of whether to offer rifabutin prophylaxis against mycobacterium avium complex to people with advanced HIV infection. It dissects part of the clinical decision-making process, of weighing up the benefits and risks of treatment against the consequences of withholding treatment, into a series of judgements about the validity and clinical relevance of the data.

The clinical problem

The incidence of *Mycobacterium avium* complex infection increases markedly in people with advanced HIV infection and low CD4 counts, with cumulative estimates of 15-24% of AIDS patients. The dissemination occurs in primary sites of infection in the gastrointestinal tract or lungs and contributes substantially to the morbidity, (causing fever, weight loss, abdominal pain and diarrhoea) and is associated with poorer survival.³ Although therapy with multiple drug combinations can improve survival and symptoms, current treatment and maintenance regimens are not ideal because of drug intolerance and a high rate of relapse and infection. Effective primary prophylaxis for patients at highest risk would seem an appropriate clinical strategy.

The evidence for prophylaxis

Rifabutin is known to have activity against *M avium* complex, both in vitro and in animal models.⁴ An early non-randomised study⁵ examined the effectiveness of rifabutin in preventing *M avium* complex in AIDS patients by comparison with a control group of AIDS patients from another institution who had not received the drug. The rifabutin-treated group had significantly better survival than the control group, but the authors were careful to point out that the survival advantage could not necessarily be attributed to rifabutin. Although the two groups were similar with respect to age and CD4 count, the proportion of patients taking *Pneumocystis carinii* pneumonia (PCP) prophylaxis, which is known to improve survival, was much higher in the rifabutin-treated group (72%) than the control group (10%). It was therefore impossible to know whether

rifabutin had affected survival or not. When designing a study to assess the efficacy of a treatment, the best way to ensure that prognostic factors (whether known, such as age, CD4 count and PCP prophylaxis, or unknown) are evenly distributed between the two groups, is to randomly allocate patients to either treatment (rifabutin) or control (placebo). Had this been done, any survival advantage in the rifabutin-treated group could have been more plausibly attributed to rifabutin. This study illustrates one of the reasons why non-randomised studies are less useful than randomised trials for estimating the effectiveness of treatment. Non-randomised studies are more likely to produce systematic differences in prognostic factors between the treatment and control groups, which can result in biased estimates of treatment effects and false conclusions. Comparison of randomised with non-randomised studies of the same treatment^{6,7} shows that the non-randomised studies tend to overestimate the effect of treatment, and may even show a beneficial treatment effect when none exists.

Is there any evidence for rifabutin prophylaxis from randomised trials? Two⁸ randomised, placebo-controlled trials have been conducted in North America. In one trial, 590 patients who had AIDS, a CD4 count of less than 200, no evidence of *M avium* complex from blood or stool culture (and satisfied additional entry criteria) were randomly assigned to treatment with either 300 mg of rifabutin daily or matching placebo. A second trial followed an identical study protocol and randomised 566 patients to either rifabutin or placebo. Patients in each treatment group were followed up approximately monthly, for an average of 185 to 231 days, and at each visit they were assessed for clinical symptoms and illnesses, and for *M avium* complex bacteraemia. Similar findings in both trials led the authors to conclude that rifabutin, given prophylactically, reduces the frequency of *M avium* complex bacteraemia in patients with AIDS and CD4 counts less than 200. Commenting on these findings in the same issue of the *New England Journal of Medicine*, a special report from the Public Health Service Task Force on prophylaxis and therapy for *M avium* complex⁹ recommended that patients with HIV infection and a CD4 count of less than 100 should receive prophylaxis against *M avium* complex. Opinions and practice in the UK vary considerably. Using the criteria

Department of Sexually Transmitted Diseases, Division of Pathology and Infectious Diseases, University College London Medical School, and Camden and Islington Community Health Services NHS Trust
J M Stephenson
I G Williams

The Mortimer Market Centre, off Copper Street, London WC1E 6AU, UK

Address correspondence to: Dr J M Stephenson.

Accepted for publication 22 June 1996

Table 1 Readers guides for an article about therapy or prevention

Are the results of the study valid?*Primary guides:*

- Was the assignment of patients to treatments randomized?
- Were all patients who entered the trial accounted for adequately at the end?
- Were patients analysed in the groups to which they were randomised?

Secondary guides:

- Were patients, health care workers/researchers "blind" to the treatment allocation?
- Were the groups similar at the start of the trial?
- Aside from the experimental intervention, were the groups treated equally during the trial?

What is the net impact of treatment?

- How important and how large were the treatment effects?
- How precise was the estimate of the treatment effect?
- How great are the likely benefits in relation to the potential harms?

Adapted from Guyatt *et al*, *JAMA* 1993;270:2598-601.

summarised in table 1, we can assess whether the results of the trials were valid and clinically important.

How valid are the results of the study?

Were patients randomly allocated to the treatment or control groups? In both trials,⁸ a randomised schedule was used to allocate patients to either rifabutin or placebo. This satisfies the first criteria for assessing validity.

Were all patients who entered the trial accounted for adequately at the end? We then need to consider whether all the patients who entered the trial were accounted for at the end. The aim at the conclusion of a trial is to know the outcome of every patient who entered it. If a substantial number of patients are unaccounted for (lost to follow-up) or not assessed for the outcomes of interest (such as disease, symptoms, treatment side effects) then the validity of the trial is threatened. Patients who are lost to follow-up often fare differently from those who reach the end of the trial. Failure to attend follow-up visits may be related to adverse effects (including death) or it may mean that the patient feels well and has little incentive to attend the clinic. In short, important bias can occur if a substantial proportion of patients are lost to follow-up. Whether this bias over or under estimates any treatment effect depends on whether the loss to follow-up was related to positive or negative events, and how this was balanced between the treatment and control groups. In the trials of rifabutin prophylaxis, only a small proportion of patients were lost to follow-up, and the proportions were similar in each group (7% of the placebo group and 6% of the rifabutin group). Thus the second criterion for assessing validity (table 1) is satisfied.

Were patients analysed in the groups to which they were randomised? We also need to ask whether patients were analysed in the groups to which they were randomised, regardless of whether they actually took the treatment as prescribed or not (that is, whether an "intention-to-treat" analysis was done). There are two important reasons for analysing the data according to intention to treat. First, patients often omit treatment for reasons related to

prognosis, and second, there will always be non-compliant patients in clinical practice. Excluding patients who did not take their treatment from the analysis would therefore destroy the unbiased comparison provided by the randomisation (particularly if patients were less compliant with one of the treatments than the other), and fail to represent the "real-life" situation in clinical practice. In the rifabutin trials, the researchers did analyse the data by intention to treat, thereby satisfying the third criterion for validity. Since the three most important criteria for validity have been satisfied, we can go on to look at some additional criteria (table 1).

Were patients and health workers/researchers "blind" to the treatment allocation? Both patients and researchers/health care workers are likely to have opinions or expectations about the effect of a new treatment. Equally, they will not expect a placebo to exert clinical effects. Knowing whether a person is on active treatment or placebo may therefore influence the occurrence (through the patient's knowledge) or assessment (through the researcher's knowledge) of the trial outcomes. "Blinding" patients and researchers to the type of treatment taken, so that neither group knows whether a patient is on active treatment or placebo, is a highly effective way of preventing such knowledge from biasing the outcome of the trial.

In some situations, as in a trial of surgery versus radiotherapy for malignant disease, it is impossible to blind the patients or the health care worker. It may, however, be possible to blind the person who assesses the outcome measures, for example the radiologist assessing metastatic bone disease or the pathologist examining biopsy material. Wherever possible, those who assess important outcomes or trial end-points should be blind to the group assignment.

Were the groups similar at the start of the trial? The greater the similarity between actively treated and control groups, for all factors that affect outcome except the treatment itself, the greater the likelihood that differences in outcome are due to the active treatment. One of the most effective ways to achieve similarity in prognostic factors is exclude patients with certain characteristics from entering the trial, such as smokers. Another way is to randomise large numbers of patients to one or other group, and this is the only way to balance unknown prognostic factors. Randomisation can result, by chance, in an uneven distribution of prognostic factors between treatment groups, but as the sample size (number of patients recruited) increases, the chance of this happening becomes smaller and smaller. If the treatment groups are not similar at baseline, the study is not necessarily invalid. Statistical techniques can be used to take account of baseline differences and adjust the study results accordingly. Validity is stronger, however, if the adjusted and unadjusted results both lead to the same conclusion.

Table 2 Incidence and relative risk of clinical manifestations of disseminated *M avium* complex in both studies combined

Clinical manifestations	Incidence		Relative risk*	p value
	Placebo (N = 580)	Rifabutin (N = 566)		
	No (%) of patients			
Fever	213 (37)	166 (29)	0.74	0.003
Fatigue	221 (38)	180 (32)	0.76	0.006
Karnofsky score†	203 (35)	168 (30)	0.79	0.026
Anaemia	309 (53)	263 (46)	0.80	0.008
Alkaline phosphatase elevation	155 (27)	116 (20)	0.68	0.002
Hospitalisation	218 (38)	180 (32)	0.81	0.035
Weight loss	286 (49)	267 (47)	0.91	0.257
Night sweats	94 (16)	77 (14)	0.81	0.179
Abdominal pain	99 (17)	102 (18)	1.04	0.745
Diarrhoea	121 (21)	125 (22)	1.05	0.699
Therapy withdrawn	(8)	(16)		

*Here, the relative risk is not simply the proportion of patients developing the end-point in the rifabutin group divided by the proportion in the placebo group. This is because the analysis also takes into account the time to development of each end-point.

†Decrease of > 20% in Karnofsky score.

Were the groups treated equally during the trial?

We can also assess the extent of similarity between the treatment and control groups by asking whether, except for the treatment itself, both groups were treated equally during the conduct of the trial. Double-blinding should ensure that this is so. In non-blinded trials, it is particularly important that patients in each group have the same opportunity for the outcome of interest to be diagnosed or reported. Less frequent questioning or assessment of a control group compared with an actively treated group, for example, might result in a falsely low disease incidence in the controls and thereby underestimate the true treatment effect.

In the rifabutin trials, markers of disease stage and rate of progression, such as CD4 count, time between AIDS diagnosis and randomisation, and age, were well balanced between the treatment and control groups. The trial was double-blind and all patients were assessed at monthly intervals for the outcomes of interest. From this brief appraisal, we can conclude that the trials of rifabutin used sound methods to reduce bias and therefore provide a valid basis from which to consider the clinical relevance of the findings.

What is the net impact of treatment?

If a valid study of an intervention clearly shows that patients receiving the intervention had a worse prognosis or poorer survival than the control group, that intervention can be discounted from further use. If, more commonly, the study shows some beneficial effect(s) from treatment, the size and importance of the benefits need to be compared with the size and importance of any adverse effects (including cost) before the treatment can be judged to be worthwhile. In addition, the applicability of the findings to patients outside the trial needs to be considered.

How important and how large were the treatment effects? The importance of treatment effects is reflected in the type of end-points or outcome measures used in a trial. Measures of clinical morbidity (for example AIDS-defining illnesses) or mortality are more useful than so-called surrogate markers or proxy measures

(such as CD4 count below 200) which are of less importance to patients, and not always good predictors of disease. In the rifabutin trials, the primary end-point was the development of *M avium* bacteraemia—a microbiological end-point that may or may not cause morbidity. Secondary end-points (table 2) were clinical symptoms associated with *M avium* complex, including fever, fatigue, night sweats, diarrhoea and abdominal pain. The authors commented that an improvement in clinical outcomes would indicate that rifabutin had prevented infection in the body, rather than merely inhibiting the growth of *M avium* complex in blood cultures. Survival was also compared between the rifabutin-treated and placebo groups.

The results showed that rifabutin significantly reduced the incidence of *M avium* bacteraemia from 17% (placebo group) to 8% (rifabutin group) in one trial, and from 18% to 9% respectively in the other trial. Statistically significant reductions (table 2) were also found in fever and fatigue, but the frequency of weight loss, night sweats, abdominal pain and diarrhoea were all very similar in the two groups. The death rate was non-significantly lower (5.8%) in the rifabutin group than in the placebo group (8.1%).

There are various ways of expressing the size of a treatment effect (table 3). One is simply the difference in frequency (risk difference) between the treatment and control groups. For example, the risk difference for *M avium* bacteraemia in one trial was 17% minus 8% = 9%. Alternatively, the treatment effect can be expressed as a risk ratio—in this example 8%/17% = 0.47. Often, the risk ratio is expressed as a percentage of the risk in the control group (the relative risk reduction). In this example, the relative risk reduction was $(1 - 0.47) \times 100\% = 53\%$. In other words, the risk of *M avium* bacteraemia in the rifabutin group was 53% lower than the risk in the control group. In a trial where there is no treatment effect at all, the risk difference and the relative risk reduction will be 0%, and the risk ratio will be 1.

Taking one of the clinical end-points, fatigue, the size of the treatment effect can be expressed as a risk difference of 6%, or a risk ratio of 0.84 or a relative risk reduction of 16%. (In the paper, the calculation of the risk ratio takes into account the time to development of fatigue. Since the time-dependent risk ratio is lower (0.76) than 0.84, we can infer that patients on placebo developed fatigue more quickly than those on rifabutin). Since there was no treatment effect on some of the more distressing symptoms associated with *M avium* complex, we can judge that rifabutin had less effect on clinical outcomes than on the bacteriological end-point. There was a tendency to better survival in the rifabutin group (and this effect may have been underestimated because some of the patients in the placebo group ended up taking rifabutin), but the trial was not large enough to exclude the possibility that the survival difference had occurred by chance (see below).

Table 3 Measures of beneficial and harmful effects

	General case	<i>M avium</i> bacteraemia
Risk of disease in treatment group 1 (control)	$X_1\%$	17%
Risk of disease in treatment group 2 (active treatment)	$X_2\%$	8%
Risk difference (absolute risk)	$X_1 - X_2\%$	9%
Risk ratio (relative risk)	$X_2 \div X_1$	0.47
Relative risk reduction	$\frac{X_1 - X_2}{X_1} \times 100\%$	53%
Number needed to treat (NNT)	$\frac{1}{X_1 - X_2} \times 100$	11
Risk of harm on treatment, for example Risk of discontinuing rifabutin = 16%	$Y\%$	
Number of patients harmed for each one who benefits	$Y\% \times \text{NNT}$	$16\% \times 11 = 1.8$

How precise was the estimate of treatment effect? Having considered the size and importance of the treatment effects, we need also to ask how precisely they were measured. The precision matters because the treatment effect measured in a trial is only an estimate of the true effect. If a particular randomised trial was carried out twice, using an identical protocol, one would not expect to get exactly the same answer (treatment effect) in each study. Even with the most valid study design possible, one would expect the size of the treatment effect to vary somewhat between the two trials, purely by chance. As mentioned above, the effect of rifabutin was measured in two trials following an identical protocol and the relative risk reduction was 53% in one and 50% in the other. However, if one trial had recruited a thousand patients, and the other only fifty, one would intuitively place more confidence in the result from the larger trial. This is because the probability that the estimated treatment effect varied widely (by chance) from the true treatment effect would be much lower for the large trial than for the small trial. In other words, the precision of the estimated treatment effect would be much greater in the large trial.

The degree of precision is reflected by the width of the statistical confidence interval around the estimated treatment effect.¹⁰ For example, the confidence interval around a 53% relative risk reduction in a trial of a thousand patients might be 48% to 58%, compared with an interval of 23% to 83% in a trial of fifty patients. When, as conventionally, the interval is the “95%” confidence interval, it defines the range of values that includes the true relative risk reduction 95% of the time. (In probability terms, if the same trial had been repeated 100 times, the estimated effect in 95 of those trials would lie within the 95% confidence interval).

In describing the different ways of expressing a treatment effect (see above), we noted that when there is no treatment effect at all, the risk difference and the relative risk reduction are both zero and the risk ratio is one. It is more likely in a small trial (than in a large trial) that wide confidence limits around the risk ratio will include one, making it hard to exclude no treatment effect. The same conclusion would be drawn if the confidence interval

around the risk difference, or the relative risk reduction, included zero. If the 95% confidence interval includes the possibility of no treatment effect (that is, if it overlaps one or zero respectively) then the treatment effect is *not* statistically significant at conventional levels, that is, the p value would be greater than 0.05. However, the confidence interval provides more information than the p value, because it indicates how precisely the treatment effect has been measured and not simply whether it is “statistically significant”, yes or no.¹⁰

No confidence intervals were presented in the study of rifabutin. However, the 95% confidence interval around the risk difference (9%) in one of the trials can be calculated as 3.7% to 14.3%. Since zero lies outside the lower limit, the possibility of a treatment effect being found in the trial when none truly exists is small. The confidence interval is moderately wide, reflecting the moderate size of each trial and resulting degree of precision in the estimated treatment effect.

How great are the likely benefits in relation to the potential harms? It is clearly important that the frequency and severity of adverse effects are reported as well as any beneficial effects. Chemotherapy or antiretroviral drugs which prolong survival are of limited value if they substantially worsen quality of life. In the rifabutin trials, the incidence of adverse effects was similar in the rifabutin (51%) and placebo (50%) groups. However, the proportion of patients who had to stop treatment because of side effects was twice as high in the rifabutin group (16%) as in the placebo group (8%). The reasons given for stopping treatment were rash, gastrointestinal intolerance of the drug and neutropenia.

The net impact of the likely benefits and harms can best be considered by calculating the number of patients you need to treat in order to benefit one individual, in relation to the number of treated patients who are likely to suffer harm. Returning to the rifabutin trial, where the difference in risk of *M avium* bacteraemia between the rifabutin and control groups was 9%, we can say that if 100 persons, like those in the trial, took rifabutin for around 200 days, bacteraemia would be prevented in 9 of them. Therefore, we would need to treat $100/9$ (the reciprocal of the risk difference) = 11 persons, to prevent one of them from getting bacteraemia. This concept has become known as the “number needed to treat” (NNT). If we take one of the clinical benefits, such as reduction in fatigue, we can calculate that 17 patients (reciprocal of the risk difference, 6%) would need to be treated to prevent one from getting fatigue. On first consideration, patients and health care workers might consider it worthwhile to offer rifabutin on that basis. However, since the probability of having to stop rifabutin because of side effects is 16%, $17 \times 16\% = 2.7$ of those treated patients would suffer intolerable side effects for every one who benefited. The net impact of rifabutin prophylaxis may therefore cause

more morbidity than it prevents. In such a situation, it would be important to know about potential long term benefits, such as improved survival. Larger randomised trials than the two described here would be needed to determine whether rifabutin prophylaxis improves survival in people with advanced HIV disease.

Other considerations

In deciding the relevance of trial findings for routine clinical practice other factors may need to be taken into account. Rifabutin and other rifamycin derivatives are an important component of drug regimens for treatment of mycobacterial disease. The widespread use of rifabutin as primary prophylaxis in patients also at high risk from tuberculosis raises concern about the possibility of an increasing incidence of drug resistance in patients presenting with active tuberculosis. Transient mycobacterium complex bacteraemia can occur, and the use of bacteraemia alone rather than bacteraemia plus clinical symptoms as an end point to the trial may have over-estimated the clinical effects of prophylaxis, although it should be acknowledged that transient bacteraemia frequently precedes the development of symptomatic disease. At the time these trials were published, there were also concerns about possible clinically relevant interactions with other drugs such as azols and anti-retrovirals, the cost of treatment and the net impact in patients with a lower overall incidence of disseminated *M avium* complex infection. For all these reasons, rifabutin for primary *M avium* complex prophylaxis has not been widely used by clinicians in the UK.

Conclusion

We have used this study of rifabutin prophylaxis against disseminated *Mycobacterium avium* complex infection in AIDS⁸ as an example of how to assess the validity and clinical importance of articles about medical intervention. It is important to examine the net impact of both benefits and harms, in qualitative and quantitative terms, before judging the value of treatment. Although the results of the study are valid, the net impact of treatment presents a weak case for rifabutin prophylaxis.

- 1 Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature II. A. Are the results of the study valid. *JAMA* 1993;270:2598-601.
- 2 Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. B. What were the results and will they help me in caring for my patients? *JAMA* 1994;27:59-63.
- 3 Jacobson MA, Hopewell PC, Yajko DM, Hadley WK, Lazarus E, Mohanty PK, et al. Natural history of disseminated *Mycobacterium avium* complex infection in AIDS. *J Infect Dis* 1991;164:994-8.
- 4 Klemens SP, Cynamon MH. In vivo activities of newer rifamycin analogs against *Mycobacterium avium* infection. *Antimicrob Agents Chemother* 1991;35:2026-30.
- 5 Nightingale, et al. Rifabutin may delay the onset of *Mycobacterium avium* complex infection (MAC) in patients with AIDS. Abstracts of the 6th International Conference on AIDS, San Francisco, June 20-24 1990;1:251 (Abstract).
- 6 Pocock SJ. The justification for randomised controlled trials. In: Pocock SJ *Clinical Trials. A Practical Approach*. Wiley, 1983;50-65.
- 7 Chalmers TC, Celano P, Sacks HS, Smith HJ. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358-61.
- 8 Nightingale SD, Cameron DW, Gordin FM, Sullan PM, Ohm GH, Chaisson RE, et al. Two controlled trials of rifabutin prophylaxis against *Mycobacterium avium* complex infection in AIDS. *N Engl J Med* 1993;329:828-30.
- 9 Masur H. Recommendations on prophylaxis and therapy for disseminated *Mycobacterium avium* complex disease in patients infected with the human immunodeficiency virus. *N Engl J Med* 1993;329:898-904.
- 10 Gardner MJ, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. *BMJ* 1986;292:746-50.